

DOCUMENT RESUME

ED 252 629

UD 024 011

AUTHOR Miller, Norman
TITLE Independent Meta-Analysis of the Effect of School Desegregation on Black Academic Achievement.
PUB DATE 34
NOTE 19p.; May not reproduce well due to light type. Paper presented at the American Psychological Association Convention (Toronto, Canada, August 24-28, 1984).
PUR TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Black Students; *Desegregation Effects; Educational Policy; Elementary Secondary Education; Interdisciplinary Approach; *Meta Analysis; Political Attitudes; *Research Methodology; Research Problems; *Research Reports; School Desegregation

ABSTRACT

This document summarizes a 1982 project of the National Institute of Education (NIE) in which seven scholars, representing a variety of disciplines and ideological positions, were commissioned to write papers on the effects of school desegregation on black achievement. As previous efforts to arrive at a consensus among experts regarding desegregation effects had been complicated by the variations in the subsets of reviewed studies, the intent of the NIE project was to develop a set of common criteria for including and excluding studies from the larger set of existing research reports so that each expert could perform a meta analysis on the same set of studies. The document discusses the problems that arose over the selection process, presents the commonalities and disparities in the findings of the experts, and discusses implications of the project for social policy development and for strategies of research synthesis. The document also presents a list of the criteria used in selection or rejection of a study, and data from the research of four of the seven scholars. (CMG)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED252629

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

✓ This document has been reproduced as
received from the person or organization
originating it.
Minor changes have been made to improve
reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Norman Miller

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

INDEPENDENT META-ANALYSIS OF THE EFFECT OF SCHOOL
DESEGREGATION ON BLACK ACADEMIC ACHIEVEMENT

Norman Miller
University of Southern California

Symposium Presentation: 1984 APA, Toronto

UD 024011

Symposium Presentation: 1984 APA, Toronto

Independent Meta-Analyses of the Effect of School
Desegregation on Black Academic Achievement
Norman Miller
University of Southern California

In 1982 the National Institute of Education commissioned a set of papers concerned with the effect of school desegregation on black academic achievement. For this purpose, seven "experts" were chosen: David Armor, Tom Cook, Robert Crain, Norman Miller, Walter Stephan, Herbert Walberg and Paul Wortman. All had previously published work on the issue. As a group they represented a variety of disciplines (education, psychology, and sociology), varied in the extent to which they had participated in desegregation litigation (from extensively to none at all), and perhaps most importantly, varied in the positions they had taken in previously published work and in their ideological and political perspective on the issue. Indeed, this latter feature was an explicit component of the procedure that resulted in their selection. Two were thought to be in favor of school desegregation, two neutral, and two opposed. Cook, a methodologist, was asked to prepare an overview of the papers written by the others.*

The intent of the project was to develop a set of common criteria for including and excluding studies from the larger set of existing research reports so that each expert could perform a meta-analysis on the same set of studies, agreed upon in advance as meeting the necessary methodological requirements. For this purpose the group was convened several times prior to their individual preparations and once afterward.

My symposium presentation will summarize this project. I will discuss the problems that arose, present the commonalities and disparities in the findings of the experts, discuss the policy implications of the results, and finally, discuss the pros and cons of this unique procedure for developing policy relevant answers from a body of scientific research.

Background: Prior Attempts to Summarize the Relation Between
Desegregation and Academic Achievement

Before further describing the project and its results, I want to set the stage by summarizing previous efforts to integrate the research on school desegregation and academic achievement. I can state in advance, however, that they have produced mixed and conflicting results.

This prior work includes the controversy in the early 1970s between Dave Armor and Tom Pettigrow published in the journal The Public Interest; the uniformly positive assessments of Meyer Weinberg, an historian with strong ideological commitment to

* These seven chapters have just recently been published by the National Institute of Education.

UD024011

desegregation, whose most recent integrative effort was commissioned by the National Institute of Education (1977); Nancy St. John's (1975) comprehensive review, which attempted to classify studies in terms of their type and strength of research design and which, despite her lifelong ideological commitment to desegregation, found no overall achievement gains; Bradley & Bradley's (1977) review, which agrees with St. John, but in addition concludes that methodological flaws impugn literally all of the studies; the chapter of Clement, Eisenhart & Wood (1976) which concludes in another review prepared for NIE that "race mixing alone has little consistent effect on black-white outcomes;" Stephan's 1978 review in the Psychological Bulletin, in which he examined fifteen published and nineteen unpublished studies and concluded that desegregation was rarely associated with a decrease in achievement but in better than 25 percent of the cases was associated with increased achievement among blacks; and finally Crain and Mahard's (1978) summary, which along with Weinberg's, reviewed larger arrays of studies than did either Stephan, St. John, or Bradley and Bradley, and interpreted more than half as showing improved scholastic achievement for black students.

To assess this literature more formally, Krol (1978) applied meta-analysis procedures to it. Basically, as all of you here undoubtedly know, meta-analysis refers to procedures for aggregating or pooling statistical outcomes of several independent studies in order to produce a single "answer." Such procedures date back at least to Thorndike (1942) and Stouffer (1948) in American Soldier. As advocated by Glass, Hunter et al., Rosenthal, and others, the preferred and typical procedure for experimental and quasi-experimental designs is to compute a standardized estimate of the effect of the experimental treatment by dividing the difference between a no-treatment and a treatment affected mean by a relevant standard deviation. These standardized effect size estimates, typically one such estimate for each study, are then averaged across studies to give a single "answer," a mean effect size. The mean "effect size" for Krol's 129 comparisons (culled from 50 studies) reveals that the academic achievement of minority children was .16 standard deviations higher as a consequence of desegregation, suggesting that desegregation, as typically implemented, has a modestly positive impact. Krol's comparison of the effects of desegregation on academic performance in the elementary grades (.22) and the effects found at the junior high and high school grades (.01) does not reach statistical significance, but it is noteworthy that for the latter grades the mean effect size is essentially zero. One wishes that he had also made a comparison of the early to later elementary grades. He also found that less benefit tends to occur in short (.11) versus longer periods of desegregation (.21), and in methodologically strong (.10) versus weaker studies (.21).

Finally, however, it is important to note three methodological points with respect to Krol's meta-analysis. First, he used a lenient definition of benefit. His procedures do not exclusively reflect minority student gains relative to those of white students; instead, they reflect in part the absolute gains of minority students, independent of what has happened to the white

students in those same school districts or classrooms. To state this another way, the benefit he finds does not reflect a closing of the achievement gap that exists nationally between the performance of blacks and whites. Whites may have benefited as much as, or more than, the blacks in desegregated schools.*

Second, and perhaps more critical from a methodological standpoint, is that Krol made no adjustment for control group selection bias, that is, differences between treatment and control groups that existed prior to desegregation. This is critical because desegregation research almost never uses designs in which children are assigned randomly to segregated and desegregated classrooms. Moreover, brighter minority students typically are over-represented in the desegregated groups. Consequently, any superiority of desegregated students (compared to segregated controls) on posttest measures may merely reflect their initial advantage.

Third, and methodologically most compromising, in studies that lacked a control group, Krol's procedures assumed that a segregated control group would have made no achievement gains during the period between the pre and post measurement of the desegregated group, an assumption that is clearly untenable. Virtually all students learn something over the course of a school year and/or improve as a result of maturation.

Crain & Mahard (1983) also used meta-analysis procedures in the last of their several efforts to assess this literature. Although, as I indicated above, they had concluded in their earlier work that more than half the studies showed improved scholastic achievement for black students, in their 1983 meta-analysis results they report an overall effect size of .065, which is negligible.

In sum then, when the current NIE project was conceived, there was little consensus among experts regarding the effect of desegregation on minority academic achievement. The major source of variation in their conclusions can be attributed to variation in the subset of studies that they review. Therefore, in the initial stages of this project it seemed important to agree on a set of criteria for identifying studies that meet some minimal standards for making a causal inference.

Procedure

In an initial meeting held between the seven panel members

* For instance, anticipating the results of my own NIE paper, among the set of studies that met the NIE panel's selection criteria, those cases for which it was possible to compare the magnitude of the gains by desegregated versus segregated whites to the comparable gain by desegregated blacks reveals that desegregated white students benefited more than three times as much as their black counterparts. This suggests that when desegregation does benefit minority students it is due to changes in schools or classrooms and not the simple consequence of desegregation per se.

and NIE staff a set of criteria was developed for rejecting desegregation studies. This was undertaken as an abstract task without specific reference to the details of any particular study. Application of the agreed upon criteria resulted in selection of 19 studies. (Parenthetically we were all indebted to Paul Wortman and his collaborators for assembling the total population of 105 studies and for the actual culling of the 19 "acceptable" studies according to the guidelines presented in Table 1.) I'm sure that no one expected that application of the criteria would result in such a small final set. Consequently, the fact that only 19 met our criteria bears further comment.

Insert Table 1 about here

First, it should be apparent that most of the original 105 studies are unpublished and form a renegade literature, existing as School of Education doctoral dissertations, mimeographed school district evaluation unit reports, etc. As noted previously, virtually none are true experiments in the sense that students were randomly assigned to treatment groups. Many lack control groups making it impossible to assess the extent to which gains are maturational or due to desegregation. Many are cross-sectional studies which means that differences, when found, may merely reflect differences that existed between the comparison groups prior to desegregation.

Among the criteria adopted, the panel's requirement that there be pretest and posttest means and standard deviations is probably the most important single requirement in that it permits some adjustment for the fact that the comparison groups typically do differ prior to desegregation.

At this point one might conclude that except for variation caused by minor (or major) mistakes in calculation the six panelists would essentially come to the same numerical outcomes. This is not necessarily the case, however. First, two of the six did not choose to perform a meta-analysis on the 19 core studies. Bob Crain preferred to base his review on the outcomes for the 93 studies submitted to meta-analysis by Crain & Mahard (1983). In his review he took the position advocated initially by Glass, namely, that basing a summary on all the literature is likely to be superior to a strategy that is selective. This view however, rests on the assumption that biasing effects are randomly distributed, or, in Don Campbell's language, that there is a "heterogeneity of irrelevancies." Within the school desegregation literature, however, the biasing effects more often than not seem to be systematic rather than random. Moreover, Crain eventually based his best estimate of effect not on the entire array of studies, but instead, on a subset quite different than that used by the panel. Herb Walberg organized his review around the comparison of the effects of school desegregation (as defined by the meta-analyses of other panelists) to meta-analysis outcomes for other educational programs designed to improve academic performance and argued that other reforms or programs are more effective in this regard than is desegregation.

Inspection of Table 2 reveals some variation among the four

remaining panelists in their pooled mean effect size for verbal and math achievement.

Insert Table 2 about here

Inspection of the effect sizes that panelists computed for individual studies exhibits considerably more variation (see Miller and Carlson's Table 2).

Insert Table 3 about here

The mean effect size variations seen in Table 2 reflect in part the fact that no panelist included all 19 studies and the four differed in the number and/or cases they excluded. Additionally, however, panelists differed in (1) whether or not they counted comparisons at different grade levels within a single study as distinct cases, or instead considered each study to be the unit of analysis; (2) their selection of control groups within individual studies; (3) the number and types of tests they included in effect size computations; (4) whether or not adjustment was made for fan spread and if so, how; (5) whether or not effect size estimates were corrected for the unreliability of the achievement tests; and other details of procedure.

On the other hand, all four panelists used a formula for computing effect sizes that did correct for pretreatment and differences between experimental and control groups. The impact of such corrections varied, of course, as a consequence of various other differences in procedure, some of which I've just mentioned. Additionally, no panelist weighted effect-size estimates by sample size. Although Hunter et al (1983) advocate such weighting, I believe that it is only appropriate when variation among effect size estimates merely reflects sampling error and not the additional effect of moderator variables. In the present case, as developed in my own NIE paper (co-authored with Michael Carlson), it is clear that the variation in effect size estimates across studies exceeds that which is to be expected from sampling variability. (Were such weighting applied to the pooled results of the four panelists, their overall outcomes would differ dramatically and would all be close to zero, or even negative, because the sample size for the Sheehan and Marcus study, which yields negative effect size estimates, exceeds 2000 whereas the sample size for many of the studies yielding positive effect sizes are rather small, e.g. $N < 50$.)

Despite the variation among panelists in their procedures, there is rather remarkable overall convergence in their final outcomes. The correlations across studies between my own effect size estimates and those of Armor, Stephan, and Wortman are + .87, + .76 and + .74 respectively. The average effect size of the panelists values is .11, only trivially higher than the value of .10 that Krol obtained for his "better" desegregation studies, and also comparable to the value that Crain & Mahard (1983) obtained when their results for studies with (1) random subject assignment; (2) control groups of segregated black children; and (3) pretest-posttest measurement were aggregated across dependent measures. Parenthetically, as I indicated earlier, Crain re-

jects this value in his own paper and prefers instead an estimate based on studies where desegregation begins in kindergarten or first grade. Although in principle there are defensible arguments for his position, many would judge that the methodological problems that arise seriously compromise the specific outcomes. In particular his effect size estimates typically lack pretest scores that allow correction for pretest selection differences; where a pretest score is available it is often an ability (IQ) rather than an achievement measure, thereby compromising its appropriateness; and finally, (pretest) scores at kindergarten level are more unreliable than those at higher grades.

As an esoteric exercise in partitioning and accounting for the variation that does appear in the results for individual panelists, I present Miller & Carlson's Table 4, which seeks to account for the disparity between our own results and those of Armor, whose outcome was most negative, and differed most from our own. As you can see, we can account for all but .003 of the + .082 value by which our own mean effect size exceeds his.

Insert Table 4 about here

Tom Cook, whose paper comments on the papers of the other panelists, presents a summary of the conclusions that emerge when the panelists' results are taken at face value.

Insert Table 5 about here

In his own paper, however, Cook is reluctant to take the panel's pooled effect size estimates at face value. He finds that median and modal values diverge from central tendency estimates based on the mean, reflecting, as he sees it, substantial skew in the distribution of effect size estimates.

Insert Table 6 about here

Thus, Cook's own conclusions diverge from those based on the pooled outcome of the panel with respect to the reading gains, emphasizing that whether or not there is an effect depends on the measure of central tendency employed. The implication of the lower estimates for the median and mode is that "the mean differences are found, not so much because the 'average' effect of desegregation on reading is positive but because -- in the panel's sample at least -- some school districts made atypically large reading gains that skewed the distribution of effect sizes." Furthermore, those studies that did produce large absolute effect size estimates had small sample sizes.

Implications for Social Policy

Cook then goes on to analyze the meaning of the "face-value pooled results of the panel," namely, the absence of gains in math and the presence of 2 to 6 week gains in reading. He does this by separately considering the perspectives of the following stakeholders: (1) protagonists of school desegregation; (2) antagonists of school desegregation; (3) persons planning desegregation activities; and (4) persons honestly seeking to learn what

school desegregation has accomplished. To some degree there is a correspondence between the policy implication drawn by particular individual panelists and those Cook infers for the different stakeholder groups.

In my own NIE paper, much like Cook's hypothetical protagonists, I argue that the legal imperative for school desegregation was to reduce stigmatization of blacks. Termination of segregated schooling reduces stigmatization, at least at a perceptual level, and programs to improve intergroup relations combats it in a positive sense. Schools cannot readily implement programs for reducing prejudice and increasing intergroup acceptance if children from different racial-ethnic groups do not attend the same schools. Therefore, if desegregation has no adverse effects on academic achievement, desegregation should be promoted (within the constraints that sufficient money and resources be made available for implementing programs designed to increase intergroup acceptance).

Antagonists can, on the other hand, as does Armor, declare the academic gains to be trivial or zero, emphasizing what has happened in school districts rather than what might happen.

With respect to persons whose job or position requires that they plan desegregation programs or activities, the current analysis by itself provides little that would add to one's confidence about being able to put together a desegregation plan that would raise reading or math achievement. Perhaps the exception to this lies in the hints from Walberg's paper regarding programs apart from desegregation that by themselves are likely to be academically productive. How these would interact with desegregation plans, of course, remains unknown.

Finally, with respect to those who are honestly seeking to learn what desegregation has accomplished, Cook states that the panel's papers will be of help "by questioning the utility of so global a label as 'desegregation.' Miller's analysis shows that after the mean effect size is accounted for, more variance remains than is due to chance. This suggests that systematic forces have to be taken into account over and above whether desegregation took place if there is to be any reasonable prediction of effect sizes."

Another group that Cook might have included among the stakeholders is researchers who are interested in school desegregation. For this group many worthwhile research ideas are suggested, either explicitly or implicitly, in the various papers. I will take time to mention only two. (1) Stephan suggests that the positive effect found for reading (but not math) reflects the greater exposure to white (as opposed to black) language, grammar, syntax and/or vocabulary. Miller & Carlson provide additional support for this notion by showing that reading gains are greater in desegregation plans involving smaller percentages of blacks in the desegregated classroom; smaller percentages imply greater exposure to white language. (2) Several panelists, while recognizing the hazards of searching for moderator variables when there are so few cases or studies, note that effects

are more positive for voluntary plans. In addition, Cook notes that effect sizes are also more positive when pretest differences are greater. Is the latter a methodological artifact or instead does it add further strength to the positive post decisional consequences of choice as inferred from the implied selection differences that characterize desegregated as opposed to segregated groups?

Implications Regarding Strategies of Research Synthesis

This project is, to my knowledge, unique in its approach to knowledge synthesis, drawing as it does upon experts with presumably different commitments. In previously published work, Crain and Wortman had concluded desegregation improves performance; Armor & Miller had concluded it didn't; and Walberg and Stephan had taken more neutral stands. A convergent estimate of effect emerging from such a heterogeneous group should have considerable policy impact. There is in fact, however, some correspondence between panelists' combined verbal-math estimate and their presumed prior position. Wortman (.17) and Crain (.30 -- based on the subset of studies that he believes best represent the truth, namely, those on kindergarten and first grade students that use randomized designs) produced the highest values. Armor produced the lowest (.04). Stephan (.14) was intermediate. Least fitting expectations was Miller (.12) who also was intermediate. A more careful or thorough analysis of panelists ideological and political alignments would, I believe, add strength to the view that ideology influences outcome. For instance, Walberg concluded that desegregation holds relatively little promise as an educational reform; though selected as a "neutral", the fact that he has consistently given court testimony against desegregation may bespeak a more anti view than assumed. Though Miller too, has consistently testified against desegregation, he is not ideologically antagonistic to it; instead he opposes mandatory plans that emphasize desegregation per se and lack meaningful augmentation of educational and intergroup relations programs. Thus he is probably more "intermediate" than "anti." Nevertheless, the convergence of overall estimates among meta-analyses performed on the panel's core set of studies is, as I indicated earlier, quite striking. And as Cook notes, the divergence among panelists appears more clearly in the policy implications they draw from their essentially common outcomes.

Another good point Cook makes is that selection of a common set of criteria for discarding studies served to minimize the impact of ideology upon computational outcomes. This too appears "confirmed" in that the two panelists who choose to analyze sets of studies that deviated substantially from the panel's core set, Crain and Wortman, produce the most deviant mean effect-size estimates as their preferred "answer."

A number of other points warrant some discussion. First, the panel's composition emphasized persons expert in the relevant content area, desegregation. Alternatively, it could have emphasized methodology (e.g. six methodologists and one desegregation expert). I favor the former over the latter, primarily because content specialists seem more likely to be aware of the special

problems idiosyncratic to the particular research area. For instance, even though randomized experiments are ordinarily far more trustworthy than quasi-experimental designs, general methodologists may not be aware that in the desegregation literature, (1) most of that small subset of studies which do use random assignment, even though authored by persons with different institutional affiliations, are based on Project Concern data from Hartford and consequently are not in fact independent from each other, and (2) in desegregation studies so-called randomized designs curiously seem to exhibit pretest differences that question the validity of the random assignment (e.g. Gerard & Miller, 1975). Or, taking another example, content specialists may be more cognizant of methodological concerns that are particular to their area such as, for instance, the importance of considering "fan-spread" problems in educational research.

Another issue is the explicit attempt to construct a panel with divergent views. Although Wortman (Wortman & Bryant, 1984) argues in favor of a neutral panel, and although neutral panels apparently do reach a closer consensus, divergence is more likely to elicit arguments and concerns that legitimately qualify a "single answer" conclusion. Though single answers may be desirable from the standpoint of policy construction, they misrepresent the nature of science. Moreover, they mistakenly imply that the scientific results provide a higher level of support (or a stronger rationalization) for whatever policy is formulated than they legitimately do.

A last point concerns the selection of meta-analysis as the procedure for integrating findings. Cook states that meta-analysis "depends most heavily on the assumption that the average bias is zero with respect to threats to internal, external, construct, statistical conclusion, or any other type of validity." He argues that meta-analysis is particularly problematic for cases such as the desegregation research in which biases caused by various threats to validity are systematic rather than heterogeneous in their effect. I would argue that, nevertheless, use of meta-analysis is (1) more likely to foster, if not force, a quantitative assessment of the effects of potential biases and (2) more likely to encourage the development and the application of corrective procedures which, though not infallible, yield greater validity than do qualitative reviews. Additionally, I would point out that science is to a substantial extent a social enterprise. Scientists are not immune to the social psychological phenomena that social psychologists and other social scientists discover. Social influence processes will inevitably cause certain research paradigms, experimental procedures, and operationalizations of theoretical constructs to become more popular, common, or appropriate than others. If so, the external and construct validity of meta-analysis outcomes invariably are compromised. Again, this should not argue against its use. A stupid qualitative reviewer is no better protected than is a stupid user of meta-analysis.

A second weakness that Cook notes with respect to meta-analysis is that it is "short" on its ability to provide explanation. I would argue that the ability to explain depends heavily

on the ingenuity of the individual reviewer, but, given a high level of ingenuity, the breadth of explanation extracted from a meta-analysis depends mainly upon the number of studies on which it is based and the natural variations among them in the levels of relevant contextual variables. In another presentation for this convention Carlson and I use meta-analysis in conjunction with independent judgments of the "contextual levels" of 19 relevant theoretical variables to test four theoretical models concerned with why negative affect augments helpfulness. We clearly reject two, find partial support for a third, full support for a fourth, and explain almost half of the total variance -- an explanatory feat that could never be accomplished within the context of a single experiment. For any case in which a meta-analysis comes up short on explanation, I doubt that a qualitative review can do better.

Table 1
CRITERIA FOR REJECTION OF A STUDY

- 1) Type of Study
 - a) non empirical
 - b) summary report
- 2) Location
 - a) outside USA
 - b) geographically non specific
- 3) Comparisons
 - a) not a study of achievement of desegregated Blacks (except in cases where we use a White comparison)
 - b) multi-ethnic combined
 - c) comparisons across ethnics only
 - d) heterogeneous proportions minority in desegregated condition
 - e) no control data
 - f) no pre-desegregation data
 - g) control measures not contemporaneous
 - h) excessive attrition (review must provide specific justification for the inclusion of studies with excessive attrition, but amount was not specified)
 - i) majority Black in a segregated condition (unless the reviewer provides specific justification)
 - j) varied exposure to desegregation (unless the reviewer provides a specific justification demonstrating that the variation in exposure time is not meaningful)
 - k) groups are initially non-comparable (unless the reviewer provides a specific justification that the amount of divergence is not meaningful)
- 4) Study Desegregation
 - a) cross-sectional survey
 - b) sampling procedure unknown
 - c) separate non-comparable samples at each observation
- 5) Measures
 - a) unreliable and/or unstandardized instruments
 - b) test content and/or instrument unknown
 - c) dates of administration unknown
 - d) different tests used in pretests and posttests
 - e) test of IQ or verbal ability
- 6) Data Analysis
 - a) no pretest means
 - b) no posttest means, unless the author reported pretest scores and gains
 - c) no data presented
 - d) N's not discernible

Table 2
Mean Meta-Analysis Outcomes of Four Panelists

	Verbal	Math	Pooled
Armor	.06	.01	.04
Miller	.16	.08	.12
Stephan	.15	.00	.14 (.08) ^d
Wortman	.28 ^a	.23 ^a	.17 ^c
	.26 ^b	.08 ^b	

- a. Computed with no pretest adjustments.
- b. Based on Cook's analyses of Wortman's results for the eleven of the panel's 19 studies for which Wortman made adjustments for pretest differences.
- c. Based on studies of his own choosing that had control groups.
- d. As corrected to 1 year of desegregated schooling.

BEST COPY AVAILABLE

Study	Year	Miller and Carlson (#1) ^a			Miller and Carlson (#2)			Armor		Stephan		Wortman	
		Verbal	Math	Other	Verbal	Math	Other	Verbal	Math	Verbal	Math	Verbal	Math
Anderson	1966	+ .733(3) ^b	+ .669 (1)	---	.764	+ .671(1)	---	+ .89 (1)	+ .54(1)	+ .42	+ .24	+ .95 (1)	+ .58 (1)
Baker	1967	+ .400(7)	- .203 (3)	+1.128	+ .265(7)	- .103(3)	+ .942(1)	+ .25 (2)	- .16(2)	+ .125	- .24	+ .095(1)	+ .285(2)
Bowman	1973	+ .061(1)	- .160(2)	---	+ .033(1)	- .375(2)	---	- .26 (2)	- .21(2)	+ .125	+ .055	+ .02 (1)	- .06 (1)
Carrigan	1969	---	---	---	---	---	---	+ .068(6)	---	- .04	---	- .04 (6)	---
Clark _C	1971	+ .039(2)	- .132(1)	---	d	d	---	- .01 (1)	+ .12(1)	+ .08	- .24	e	e
Evans	1973	+ .040(2)	+ .084(2)	---	d	d	---	+ .015(2)	+ .07(2)	+ .02	+ .03	e	e
Iwaniki and Gable _C	1978	- .051(3)	---	---	d	---	---	0.0 (3)	---	---	---	e	e
Klein _C	1967	+ .159(1)	+ .333(1)	+ .313(1)	d	d	d	0.0 (1)	- .08(1)	+ .23	+ .33	e	e
Laird and Weeks	1966	+ .257(3)	+ .074(3)	---	+ .343(3)	+ .127(3)	---	- .123	- .06(3)	+ .13	+ .063	e	e
Rentsch	1967	+ .202(3)	- .134(3)	---	+ .439(3)	- .225(3)	---	---	---	+ .22	- .053	+ .497(3)	- .110(3)
Savage	1971	- .080(3)	- .091(1)	+ .144(3)	- .077(3)	- .014(1)	+ .030(3)	+ .15 (1)	- .08(1)	+ .06	- .04	+ .14 (1)	- .05 (1)
Sheehan and Marcus _C	1978	- .188(2)	- .163(1)	---	d	d	---	- .16 (1)	+ .21(1)	- .07	- .08	- .16 (1)	- .16 (1)
Slone _C	1968	+ .091(2)	+ .291(1)	---	d	d	---	+ .27 (1)	+ .47(1)	+ .19	+ .22	e	e
Smith	1971	- .077(1)	+ .130(1)	---	- .101(1)	+ .256(1)	---	- .06 (1)	+ .13(1)	- .01	+ .02	- .05 (1)	+ .1' '1)
Syracuse	1966	+ .691(1)	---	---	+ .691(1)	---	---	+ .375(2)	---	+ .75	---	e	---
Thompson and Smidchens	1979	---	---	---	---	---	---	---	---	- .15	+ .04	e	e
Van Every	1969	- .166(2)	+ .543(1)	---	- .217(2)	+ .559(1)	---	- .46 (1)	+ .51 (1)	- .12	+ .14	- .44 (1)	+ .53(1)
Walberg	1971	- .049(3)	---	---	- .048(3)	---	---	- .10 (4)	---	+ .055	- .02	+ .012(4)	?
Zalep	1971	+ .671(2)	- .157(1)	---	+ .667(2)	- .157(1)	---	+ .53 (1)	- .17 (1)	+ .66	- .15	+ .65 (1)	- .15 (1)
		$\frac{N}{X} = 17$ $\frac{N}{X} = +.11^*$ $SD = .297$	$\frac{N}{X} = 14$ $\frac{N}{X} = +.077$ $SD = .284$	$\frac{N}{X} = 8$ $\frac{N}{X} = +.528$ $SD = .526$									
For Verbal and Math:		For Verbal Math and Other:											
		$N = 31$ $\bar{X} = .123^*$ $SD = .289$	$N = 34$ $\bar{X} = .159^{**}$ $SD = .326$										
$zs_1 = \frac{\bar{X}_E(post) - \bar{X}_C(post)}{\sqrt{\frac{(N_E - 1)S^2_{E(post)} + (N_C - 1)S^2_{C(post)}}{N_E + N_C - 2}}} - \frac{\bar{X}_E(pre) - \bar{X}_C(pre)}{\sqrt{\frac{(N_E - 1)S^2_{E(pre)} + (N_C - 1)S^2_{C(pre)}}{N_E + N_C - 2}}}$													
$zs_2 = \frac{(\bar{X}_E(post) - \bar{X}_E(pre)) - (\bar{X}_C(post) - \bar{X}_C(pre))}{\sqrt{\frac{(N_E - 1)S^2_{E(post)} + (N_E - 1)S^2_{E(pre)}}{N_E - 1} + \frac{(N_C - 1)S^2_{C(post)} + (N_C - 1)S^2_{C(pre)}}{N_C - 1}}}$													

Notes:

- See text section to the right for Miller & Carlson's formulas #1 and #2.
 - Numbers in parentheses are the number of effect size comparisons.
 - Uses estimates based on ANCOVA.
 - Estimates from formulas #1 and #2 are identical due to use of ANCOVAs.
 - Not pretest adjusted.
- * p<.05; ** p<.01

$$es_1 = \frac{\bar{X}_E(\text{post}) - \bar{X}_C(\text{post})}{\sqrt{\frac{(N_E - 1)S_E^2(\text{post}) + (N_C - 1)S_C^2(\text{post})}{N_E + N_C - 2}}} - \frac{\bar{X}_E(\text{pre}) - \bar{X}_C(\text{pre})}{\sqrt{\frac{(N_E - 1)S_E^2(\text{pre}) + (N_C - 1)S_C^2(\text{pre})}{N_E + N_C - 2}}}$$

$$es_2 = \frac{(\bar{X}_E(\text{post}) - \bar{X}_C(\text{pre})) - (\bar{X}_C(\text{post}) - \bar{X}_C(\text{pre}))}{\sqrt{\frac{(N_E - 1)S_E^2(\text{post}) + (N_C - 1)S_C^2(\text{post})}{N_E + N_C - 2}}}$$

E = Experimental (Desegregated) Group
C = Control (Non-Desegregated) Group

In this formula, the desegregation vs. segregation pre-post gain score difference is divided by an estimate of standard deviation that is based on the pooled posttest figures. If the pretest standard deviations tend to be low relative to those of the posttest, and if the desegregation group tends to possess a higher mean than the control group at the time of the pretest (as is the case when the fan-spread hypothesis holds), then this formula should produce larger estimates of effect size than should the first formula. This is true because the typical pretest advantage for the desegregated students, which is subtracted from the standardized posttest difference, will be weighted more heavily in determining effect size estimates.

Table 4

Analysis of Discrepancy Between Reading Effect Size
Estimates of Armor and Miller and Carlson (#1)^a

Source	Contributions
Inclusion of Reliability Correction	+ .005
Inclusion of Rentsch	- .008
Inclusion of "other" category data	+ .0358
Averaging in of extra tests excluded by Armor	+ .002
Calculational differences on same non-Ancova cases	+ .006
Calculational differences on cases where we estimated from Ancova	- .006
Difference comparison groups used in same study (Klein)	+ .0172
Armor's inclusion of Carrigan Study	+ .005
Cases within studies included only by Armor	+ .022
Total: =	+ .079
(Miller and Carlson + .159) - (Armor + .077) =	+ .082
Unaccounted difference =	+ .003

Note:

a. Table entries are based on overall means of Miller and Carlson's Verbal, Math, and "Other" tests.

Table 5

- a. Desegregation did not cause a decrease in the achievement of black children.
- b. It probably did not cause an increase in math skills, for the mean gains vary from 0 to .08 standard deviation units.
- c. It may have caused an increase in reading skills, for the mean gains vary from .06 to .26.

The range estimate for reading deserves comment, since the upper bound comes from our analysis of Wortman's eleven studies where pretest adjustments could be made. This is a considerably smaller sample than the other authors analyzed, and so should be treated as particularly tentative. Omitting it gives a revised range that permits a fourth conclusion, which I believe to be better justified than the third conclusion immediately above.

- d. The gain in reading was somewhere between .06 and .16 standard deviation units. This is between two and six weeks of gain if we follow the rule of thumb of Glass et al (1981) and associate a gain of one-tenth of a standard deviation with one month's gain in knowledge.

Table 6

Central Tendencies for Reading - 17 Common Core Studies

	Mean	Median of Comparison	Median of Studies ^e	Midpoint of Modal Category of Comparisons
Armor ^a	.13	.03	0	-.05 & +.05
Miller ^b	.16	--	.06	-.05 & +.05
Stephan ^c	.13	.07	.08	+.05
Wortman ^d	.26	.15	.04	--

^a Based on N of comparisons; Carrigan and Thompson & Smidchens omitted; Rentsch added and given Wortman values.

^b Based on N of studies; Carrigan and Thompson & Smidchens omitted.

^c Based on N of comparisons; Carrigan and Thompson & Smidchens omitted. Thus, Iwanicki & Gable and Slone added.

^d Based on N of comparisons. The sample size is considerably smaller than with other analysts, since Wortman omitted all instances where the control group standard deviation was not specifically given. This resulted in the omission of Clark, Evans, Iwanicki & Gable, Klein, Lard & Weeks, Slone, Syracuse, and Walberg, as well as Carrigan and Thompson & Smidchens. No mode was ascertainable.

^e The medians are from Miller's Table 2 for each author based on N of studies rather than comparisons.

BEST COPY AVAILABLE